

# MEDICT : A MALAY<-> ENGLISH BILINGUAL DICTIONARY FOR JAVA MOBILE PHONES

Anusuriya Devaraju

*Faculty of Information and Communication Technology,  
University Technical Malaysia Melaka (UTeM), Malaysia  
anusuriya@utem.edu.my*

Haniza Nahar

*Faculty of Information and Communication Technology,  
University Technical Malaysia Melaka (UTeM), Malaysia  
haniza@utem.edu.my*

Sivasakthi Marimuthu

*Faculty of Information and Communication Technology,  
University Technical Malaysia Melaka (UTeM), Malaysia  
sivasakti\_marimuthu@yahoo.com*

## ABSTRACT

Today, mobile phones provide number of advanced functions, among which possibility of running mobile applications seems to be the most interesting feature. It changes a standard mobile phone into a mini-size yet still powerful computer that able to run various applications, for example games, stock checkup, map guide, dictionary and so on. This paper presents a project currently in progress which aims at the construction of bi-directional Malay - English translation dictionary named MEDict. MEDict is an easy-to-use mobile dictionary that is the ideal reference tool for all those requiring quick access to word translations. MEDict will ideally suit J2ME mobile phones with GPRS connectivity. This paper delivers the initial development progress of MEDict. It covers the functional requirements and possible development approaches followed in building MEDict. The project incorporates the development of four main components: (1) a java-based mobile client that provides interface for front-end query input; (2) a back-end database for serving dictionary contents; (3) a dictionary service for handling the exchange between the mobile client's queries and the database; (4) a web system that allows the administrator to manage the dictionary contents. The key contribution of this study is the use of Java and mobile technologies to develop a translation dictionary between two languages that has not been previously supported. The developed mobile dictionary is expected to meet the needs of a wide range of users such as language learners, native speakers, travelers and business professionals in Malaysia.

## KEYWORDS

Bilingual, Dictionary, Translator, J2ME

## 1. INTRODUCTION

Dictionaries are commonly found in the form of books, electronic portable devices, CD-ROM or available online. A bilingual dictionary is a dictionary that translates words from one language to another. Bilingual dictionaries can be either mono-directional, meaning it translates only from language A to language B, or can be bi-directional, translating to and from both languages (Wikipedia). The national language of Malaysia is *Bahasa Melayu* (Malay) with English being used prevalently. As these two languages are widely spoken, bi-directional Malay - English translation dictionary is vital to avoid unnecessary difficulties when traveling, studying languages, communicating with others and in many other situations. The motivation to come up with MEDict is due to lack of Malay-English bi-directional mobile dictionary in the market compared to the amount of mobile dictionaries available in other 'world' languages such as English, French, Mandarin, and so forth.

Dictionary is an indispensable utility for every individual, but the convenience is however compromised when expected to carry paper dictionary or gadget around to have a quick access to words definitions. Besides, with paper dictionaries, user has to turn the pages of dictionary to search for a specific word. Because of the very large cost, a dictionary which is written on paper has few opportunities for revisions. Besides, the traditional dictionary can only be searched by the headword (Hideaki,1998). For CD-ROM or online dictionaries, they are not fully portable as they require computers and Internet connection to use them.

In this paper, we present possible development approaches, providing insights into the abilities of existing mobile dictionaries in Section 2. We begin Section 3 by describing the requirements and initial design of the proposed system. The conclusions are drawn in the Section 4.

## 2. RELATED WORKS

The current trend of existing works and its key features has been studied to design the new system. The following section covers several existing Java-based bilingual mobile dictionaries in the market.

*SABC Education* in partnership with the *Gauteng Economic Development Agency (GEDA)* and *Biza Telecoms (Pty) Ltd* has launched the **SABC Education Mobile Dictionary** or **MobiDic**. MobiDic allows users to send a word to a premium - rated short code number, where they will instantly receive a definition in return via SMS. It includes single word, phrase, hyphenated word and abbreviations lookups. One of the advantages of using SMS for accessing word definitions is the availability of SMS capability on all mobile handsets means high reach. However, message length supported by SMS is limited, thus a detailed word translation cannot be delivered within a single SMS to the mobile user.

**Radioamsoft** has developed various Java mobile dictionary applications such as '*I need France*', '*I need Japan*' and more. The dictionary translates the popular English words into chosen language, for example, '*I need France*' dictionary application supports the translation of English words to French language. The interesting features of these mobile dictionaries are the mobile phones do not need a particular '*language support*' to display the translation words. User can compose a phrase choosing up to 6 words by the phone menu. And the translation will be shown on the screen using the built-in small-sized 'words pictures' of selected language. As the dictionary uses built-in pictures to represent the translations, it cannot support large amount of word entries. Unlike non-Latin alphabets-based languages (Japanese, Chinese, Hindi, etc.) which require language support on mobile phone, MEDict requires no localization support for Malay language because the language is written using Latin alphabets.

**LIVE Dictionary** from *UtiSoft.com* is a translation dictionary optimized Java and Symbian compatible phones. There are number of bi-directional translations in many languages available such as Dutch, English, French, and so on. A compatibility chart for LIVE Dictionary included in website, so that users can find out which largest dictionary they can install on their mobile phones. As the whole database of words is stored in the dictionary application, the searching is possible when offline and for each word only one variant of translation is provided. However, such embedded mobile dictionary cannot support enormous data and updated regularly.

**SlovoEd for mobiles** ideally suits all phones supporting Java and connected to Internet. It support *Morphology module* which enables the translation of word in any grammatical form. The *Search history* allows quick access to last 20 searched words. The user interface is available on different European languages with cross-translation. With *Variants of the word* option, user can receive the nearest words to the translated word according to alphabet. It supports over 100 explanatory and translation mobile dictionaries available for 30 languages, except for Malay to English or vice versa translation.

## 3. PRELIMINARY DESIGN

### 3.1 Requirements Study

The study on closely related works explains how we came to choose certain methodologies or requirements to work with. The key features of the MEDict can be divided into three groups as below.

### 3.1.1 Contents

The dictionary contains possible exact word translations with usage samples, grammatical information and phonetic transcription. The word entries are multi-color formatted and arranged as follows: (i) Headwords, derived words and variants (*blue type*), (ii) Phonetic (*International Phonetic Alphabet (IPA) is used to define the pronunciations of words*), (iii) Grammar labels (*black italic*), (iv) Translations with examples (*black*). The content of MEDict is based on the *Oxford Fajar Inggeris-Melayu/Melayu-Inggeris Bilingual Dictionary*.

### 3.1.2 Easy-to-use

The user interface of MEDict dictionary is available in 2 languages- one for speakers of English and the other for speakers of Malay. The brief and clear instruction texts in the dictionary allow easy and fast navigation through words. The 'Help tool' includes explanations on abbreviations used in the dictionary and the user manuals. Apart from this, the 'Favorite Words' lets user to store up to 10 words with custom notes in the mobile phone. This persistent storage is modeled after simple record-oriented database system called *Record Management System (RMS)* from J2ME MIDP.

### 3.1.3 Sophisticated Search

'Wild Card Search' is available for looking up words with uncertain spelling. The valid wildcards are '?' represents exactly one character and the '\*' represents none or several characters. For example, if the user enters {sat\*}, all the terms that start with 'sat' will be displayed. If the user types {helium?}, the dictionary will find the word 'helium'. The 'Morphology' function assists user to search translation of the words entered in different grammatical forms like plural nouns, different tenses, adjectives and so forth. The 'Search History' allows quick access to 15 words previously looked up.

## 3.2 System Architecture

The client application is developed in J2ME for the CLDC/MIDP 2.0 profile. The dictionary service is developed using Java Server Pages (JSP) and Servlets and is hosted on Apache Tomcat web server. Communication between the server and client is achieved via HTTP protocol. This client/server architecture allows for easy extension of the functionality and is not restricted by the limited resources of the mobile device. The Figure 1 gives you a pictorial representation of the system architecture.

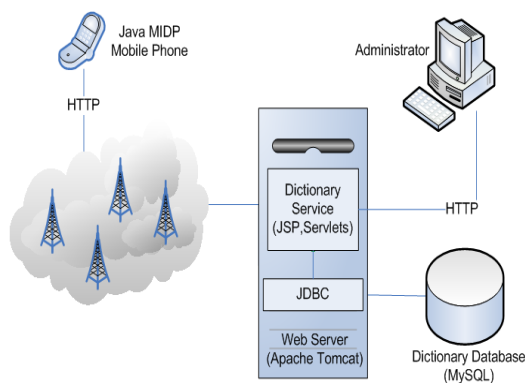


Figure 1. System architecture

### 3.2.1 MIDP Client

The client application is developed using *Sun Java Wireless Toolkit*. Other J2ME development kits are such as *Sun ONE Studio Mobile Edition*, *Nokia Developer's Suite for J2ME* and *Yospace Graphical Interface (YoGI) Toolkit*. Prior to using the dictionary, user needs to register with MEDict website. Two possible options how to install the client to the mobile: Download the installation files (\*.jad and \*.jar) through phone browsers, where it is placed in the official web site on the server or install it directly from a PC via USB cable, IrDA or Bluetooth connection.

### 3.2.2 Dictionary Service

The mobile users can insert a word in the client program and then connects to the dictionary server via GPRS service to obtain word translations. They may also use the wildcards to find words that fit any pattern. When the dictionary service received request from the MIDP client, it identifies the subscriber. Then, it retrieves the corresponding translations from the dictionary database and then sends the results to the client. If the

requested word not existed in the database, the service uses 'Best-matched Search' option to retrieve a list of the nearest words for the word user entered. Then, it delivers the list to the user so that he can perform a new search. The next actions the subscriber is likely to ask for a translation of different word or save the search results on his mobile phone via the 'Favorite Words' option.

The following Figure 2 shows the MEDict mobile client application program flow. As the 'favorite words' are stored in the dictionary client application, the searching of these words with their translations is possible when offline.

### 3.2.3 Dictionary Database

Unlike embedded mobile dictionary, the advantages of placing the dictionary database on the server is it can support enormous data and updated regularly through web system. The structure of the paper-based bilingual dictionary has been studied and customized into a database design. The dictionary database is created using MySQL database. Java Database Connectivity (JDBC) is an API that provides access to dictionary database from the Java code.

### 3.2.4 Dictionary Content Management

Administrator can use the web system to add, edit or delete word entries from the dictionary database. This will make the mobile dictionary up-to-date with ability to include the most recently published words and their meanings.

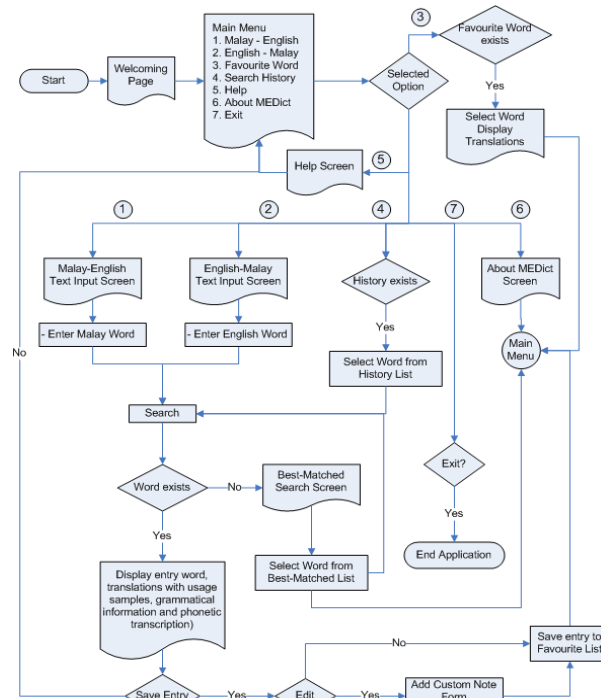


Figure 2. MEDict client application program flow

## 4. CONCLUSION

MEDict can offer great potential benefits for users: it is quick and easy to use and provides unlimited number of translations lookup's 24 hours a day without any restrictions. It benefits all, who do not want to carry a bilingual Malay<-> English dictionary everywhere but just a mobile phone. It can assist users in improving language skills and enhancing reading and writing abilities in both languages. To sum up, this project will be the beginning of more initiatives for developing Malay<-> English mobile-based dictionaries in Malaysia.

## REFERENCES

- Hideaki Ozawa, et al., 1988, DIS: A User Interface System design for the dictionary's database. *IEEE Workshop on Languages for Automation: Symbiotic and Intelligent Robots*, College Park, MD, pp. 164 – 169.
- International Phonetic Association, *International Phonetic Alphabet (IPA)*, <http://www.arts.gla.ac.uk/ipa/ipa.html>
- jDictionary Mobile, *Advanced English Dictionary (AED)*, <http://jdictionary-mobile.com/technology.html>
- Paragon Software Group, *SlovoEd for mobiles*, <http://java-software.penreader.com/SlovoEd.html>
- RadioamSoft, <http://www.radioamsoft.com>
- SABC Mobile, *Mobile Dicitonary – MobiDic*, <http://www.sabcmobile.co.za/mobidic.htm>
- Sun Java Wireless Toolkit, <http://java.sun.com/products/sjwtoolkit/>
- UtiSoft.com, *LIVE Dictionary*, <http://www.utisoft.com/live-dictionary/for-mobile-phone/>
- Wikipedia - the free encyclopedia, [http://en.wikipedia.org/wiki/Main\\_Page](http://en.wikipedia.org/wiki/Main_Page)